Final Assignment:
Predictive Analytics (KAN-CDSCO1005U)

Evaluating Global Surface Temperature Anomalies: A Comparative Study of ARIMA and ETS Models in the Context of Public Perception of Climate Change

Leonard Brenk (158287)
Master in Business Administration and Data Science
Copenhagen Business School
15 June 2023

Number of characters: 21,682
Number of pages: 10

Contents

1 Introduction

During the last decade, climate change has shown to be a highly discussed and controversial topic. While some people believe to be part of the last generation on this planet, others deny climate change altogether. As topics of such high political and public relevance can be misleading in many cases based on individual interests and influences, scientific findings are essential to prove what is actually true. In the case of climate change, often the absolute temperatures are evaluated instead of the relative changes. In this paper, such changes are considered in a normalized and season-adjusted form, indicating how strong the trend actually is. Especially as a society's behavior is rarely based solely on facts, it is crucial to prevent myths or lies about climate change. Therefore this paper aims to identify underlying patterns of the surface temperature changes disregarding the absolute values. The dataset was created by comparing monthly data to an average value of a respective month from a baseline period from 1951-1980. Using the normalized changes from the data set retrieved from NASA, this paper evaluates the application of forecasting methods to capture the actually underlying pattern of how the temperature changes and how promising the results are if more obvious trends by absolute values are disregarded. The forecasting models ARIMA and ETS are applied and compared. The results from this study could challenge the public's perception of climate change by revealing a potentially slower and less dramatic trend, detached from the absolute temperature values. Furthermore, the results will indicate the effectiveness of certain model configurations for forecasting temperature change. Due to the multitude of influences, like data quality, data origin, model configuration and combinations, data preprocessing, considered periods, etc., the anticipated findings of this study should be interpreted cautiously.

2 About the Data

The paper assesses the estimated anomaly change of the global surface temperature using the "GISS Surface Temperature Analysis (GISTEMP)" dataset from the NASA Goddard Institute for Space Studies (GISS) (Hansen et al., 2023). The data has been updated monthly since 1880 and was collected through weather and water temperature reports as described in Lenssen et al. (2019). The dataset's analysis method and final creation are fully documented in Hansen et al. (2010). The program's baseline period used to compute the surface temperature anomalies is 1951 to 1980. The average temperature was computed for each month of the year during this period. Each month's temperature is compared to the average for that same month during the 1951 to 1980 period. As the collected data show the differences between the measured and average temperatures for a specific place and time, it is possible to create an average temperature change for the earth's entire surface. The anomalies extract the underlying pattern of the temperature's change from the actual absolute values, thus providing a scientific way to evaluate trends more accurately and to prove climate change based on a norm instead of absolute values. The data set consists of 13 columns (one for each month) and 144 rows (one for each year from 1880 until 2023). As 2023 is not complete yet, the last row has been removed. The values can reach from -0.81 to +1.37 degrees Celcius. To perform time series analysis, the shape was transformed from (143 x 13) to (1716 x 2) to have one column containing all values. In the later steps of this report, the data sequence will be shortened to increase the interpretability of the prediction. Given NASA's credibility and proven approach to data collection, there's an inherent confidence in the data's accuracy. Thanks to NASA's high-quality data collection methods, the data are complete and nicely formatted; thus, there's no requirement for further cleansing or filtering, making the data immediately usable in its provided form.

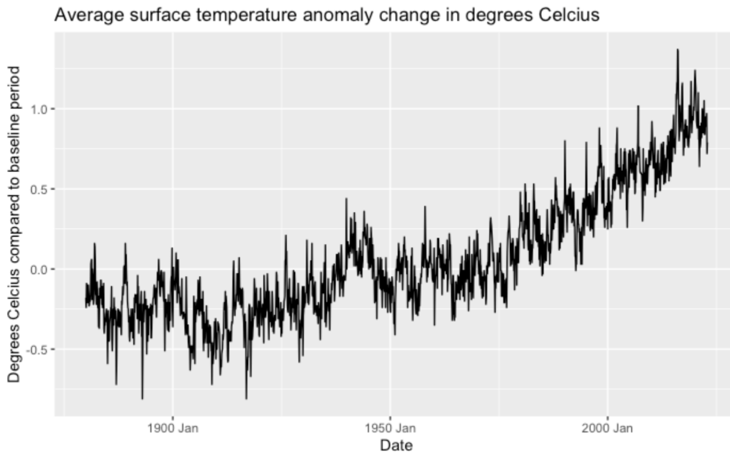

Figure 1: Plot of the entire data set

3 Exploratory Data Analysis (EDA)

As shown in Figure 1 the temperature changes are increasing most of the time except for a little decrease around 1950. Seeing that the benchmark period is in the middle of the considered values, it makes sense that the left part of the plot contains mostly negative values. At first glance, the series appears to have a clear positive trend, which is even getting stronger over time. However, the complexity of combining data from all over the world over a long period of time might lead to a notable influence on the possible trend by noise and random residuals, which will be evaluated later. The need for a possible log or box-cox transformation is assessed to achieve a clearer perspective on the more detailed development of the time series. Such can help stabilize the variance and make the graph a bit better to analyze visually. However, a Guerrero score of 0.96 shows that such a transformation would not yield much improvement in that regard.

3.1 Trend

Given the rising awareness regarding climate change in the past years, an increasing trend in temperature changes seems obvious. To accurately assess the presence of a trend, the STL decomposition is performed. It shows the components resulting in the graph in Figure 1 by dividing it into residuals, a trend, and seasonal contributions, as seen in Figure 4. The trend doesn't show a straight line but still displays an overall rise which indicates a non-stationary nature. Due to its fluctuating behavior, extracting a clear pattern, at first sight is difficult. An indication of a trend being present can be seen in Figure 2, however, which presents a very strong autocorrelation with all lags being way above the significance threshold. If one looks at the partial autocorrelation in Figure 3, it can be understood that at least 4 following lags still impact an instance, while the first carries the highest influence.

3.2 Seasonality

Intuitively, a seasonal component would be expected when dealing with temperature data, as summer and winter differ quite a lot considering the temperature. However, firstly, the data set describes anomalies, meaning the differences between the average and measured temperatures. Thus, a seasonal component

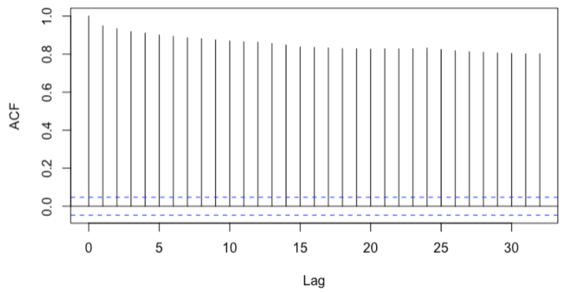

Figure 2: ACF plot of the original data

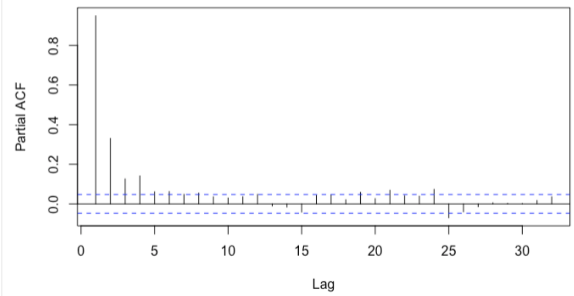

Figure 3: PACF plot of the original data

would not describe the absolute increase in temperature but the extent to which the temperature rose compared to the value in the benchmark period. Thereby, the value could be counter-intuitively higher in January than in July (while, at least in the Northern Hemisphere, July is one of the warmest months). Secondly, the global temperature varies very little in total while the regional temperatures change much more and balance each other. Figure 4 displays a very heterogenic seasonal behavior which - based on the large rectangle on the left side - doesn't seem to impact the time series development very much. Still, one could argue that a weak seasonal component is present. The amplitudes seem to build up at specific points in time and don't seem very random, while their regularity is not apparent either. When assessing Figure 5, which aims to compare the months and their relationships over the years, the only implication that can be derived is that the further back in time one looks, the more minor the changes were. The attempt to find a recurring pattern between months and years seems challenging, even when only looking at the years 2000-2022 in Appendix A.2. One might be able to detect a slight increase at the beginning of the year. That notion is confirmed in Figure 6. A seasonal behavior can be found, with the average value being the highest in March and the lowest in July, with a difference of 0.1-0.2 degrees Celcius.

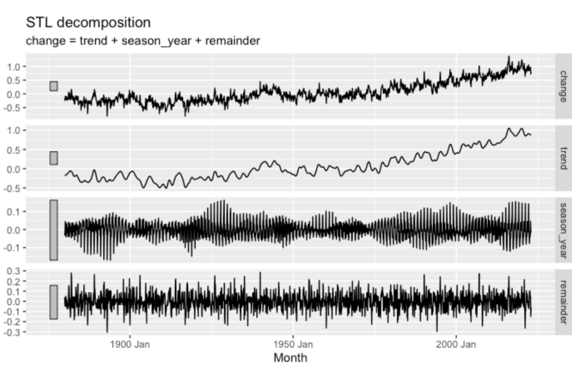

Figure 4: STL decomposition

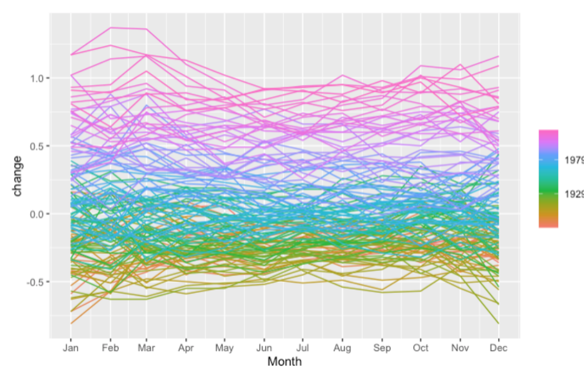

Figure 5: Temperature changes over months 1880-2022

3.3 Stationarity

Having a stationary time series or transforming it into one can be very desirable when applying specific forecasting methods. Stationarity means having a constant mean and an independent autocovariance. In

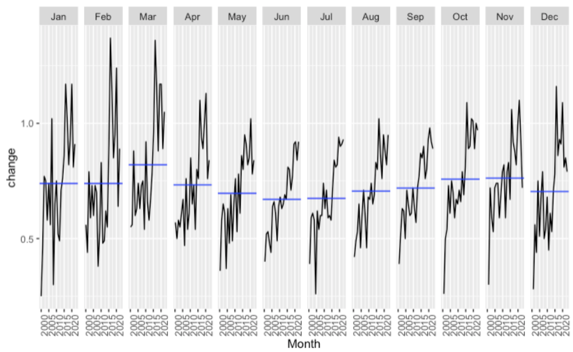

Figure 6: Temperature changes over months

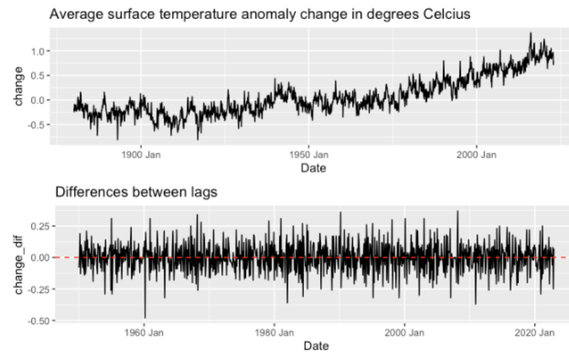

Figure 7: Original (top) vs differenced time series (bottom) with constant mean (in red)

the case of this data set, Figure 1 shows already that the time series does not have a constant mean, which is confirmed by the common truth that the climate is getting warmer. Further, the existence of a unit root is evaluated by applying the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests (Hyndman and Athanasopoulos, 2018). Having a unit root means that the time series does not tend to return to the constant average mean. With the ADF and KPSS tests, it can be found whether the time series contains a unit root or not. The threshold for significance is set to 5%. The H0 hypothesis of the ADF test suggests that the series has a unit root. While the values of the "trend" ADF (tau3, phi2, phi3) showed values smaller than the 5% value, but the other two ADF tests ("drift" and "none") presented values within the 10%+ boundary (tau2, phi1, tau1), it can be assumed that there is a unit root. The H0 hypothesis of the KPSS test is that there is no unit root; thus, the series is stationary. In order to show that there is a unit root, the H0 hypothesis should therefore be rejected, meaning the values (with tau, mu) need to be greater than the critical value of the significance threshold. The results show that this is indeed the case for both KPSS tests. More information regarding the tests and their results can be found in Appendix A.1. Having concluded that there is a unit root means that the time series is not stationary, and the variance and mean change over time. In order to transform the data into a stationary form, it needs to be differenced. The process of differencing creates a new time series that consists of the differences between consecutive values of the original series, making the series stationary and enabling the effective application of models. However, when computing the differences, one has to decide how many lags are included. To find the most effective number of lags to make the series stationary, the differencing is applied with one lag followed by another round of tests. Given the output of the tests, one might need to either do it again, thus difference twice or repeat the process using more lags. In this case, the differencing with one lag has resulted in a desirable result. All three ADF tests (tau1, tau2, tau3, phi1, phi2, phi3) showed values outside of the significance threshold, thereby rejecting the H0 hypothesis of having a unit root. The KPSS tests accepted the H0 hypothesis of not having a unit root with both values being lower than the significant values. The resulting differences are plotted in Figure 7, and the updated ACF and PACF values are shown in Figure 8 and Figure 9 below. It should be kept in mind that the significant lag in Figure 9 might be an indicator for an over-differenced time series and can influence the performance of the model.

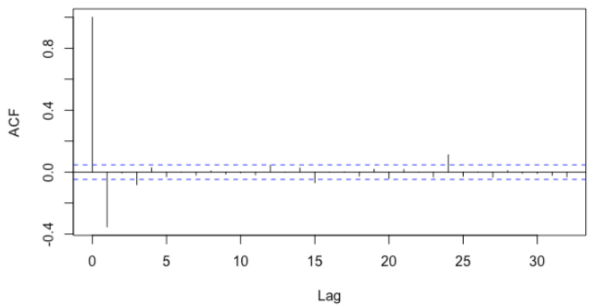

Figure 8: ACF after differencing

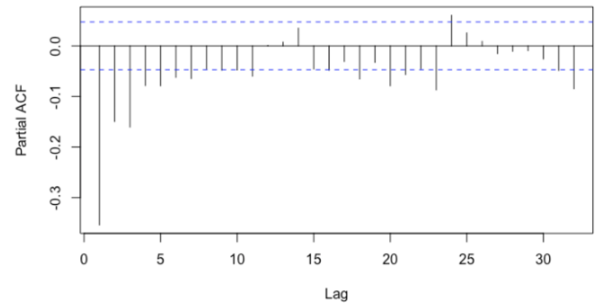

Figure 9: PACF after differencing

3.4 Structural Breaks

With the goal of forecasting values based on past observations, there must be no critical break in the data from where the training begins to where the prediction ends. If there is, one can't be sure that the prediction is accurately based on the behavioral patterns from the past and not instead on whatever happened and created a sudden change in pattern. To ensure there has not been a structural break in the time series, the Quandt Likelihood Ratio (QLR) test is performed. As the seasonality assessment did not yield a particularly clear result, the test was conducted using the first and the 12th lag. The results are shown in Figure 10 and Figure 11. As neither is crossing the red line, a structural break can be ruled out (Oswald et al., 2020).

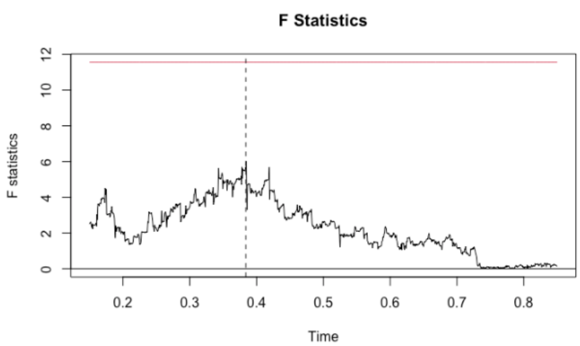

Figure 10: QLR (l1) test

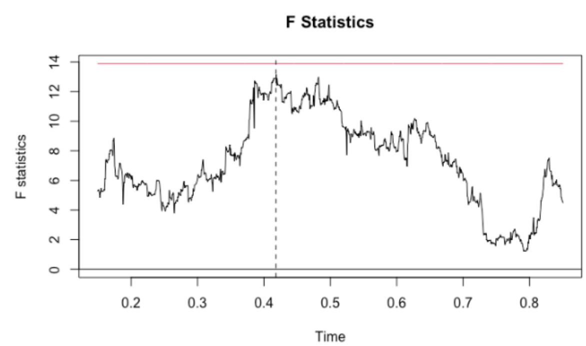

Figure 11: QLR (l1 + l12) test

4 Methodology

This paper applies and evaluates two forecasting methods, ETS and ARIMA, to predict future values. The models were picked as they are widely used methods, able to handle non-stationary time series data with trend and seasonality components. In light of the complexity of the models, the proper parameters need to be chosen using Akaike Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC), two well-known penalized-likelihood information criteria. As the data set contains 1760 instances, the data used for the forecast will be shortened to make the evaluation of the prediction outcomes more manageable and comprehensive while maintaining an 80/20 split. Using data from January 2000, the training set contains 217 instances, and the test set 59.

4.1 AutoRegressive Integrated Moving Average (ARIMA)

With its ability to incorporate trends, seasonality, autocorrelation, and non-stationarity, the ARIMA model is a widely-applied set of forecasting models. (AR) describes how strong previous observations influence a current one, (I) states how many lags are used for differencing the time series into a stationary one, and (MA) expresses the error as a combination of past error terms (Hyndman and Athanasopoulos, 2020). The model can be tuned by selecting different parameters for p (used for AR), d (used for I), and q (used for MA). In order to find such, one has to look at the ACF and PACF of the stationary time series in Figure 8 and 9. The number of significant lags will impact the choice of parameters. Looking at the PACF plot (Figure 9), there are three very significant lags right in the beginning indicating to chose p = 3. The ACF (Figure 8) shows two significant lags. However, since the second is so much less significant than the first, q = 1 seems to be the appropriate choice. As previously found, the original time series can be made stationary using 1 lag differencing, thus d = 1. When applying the automatic ARIMA model that chooses the parameters automatically, the suggested configuration is (1,1,3) - illustrating that the ACF and PACF plots can be hard to interpret perfectly. To find the best combination, the paper will evaluate the AIC and BIC score of three models: ARIMA(1,1,3), ARIMA(3,1,1), and a combination of both being ARIMA(3,1,3). As there has been detected a small seasonal pattern in section 3, each model will be tested with a seasonal component of (0,0,0) and (1,0,1). As the seasonal component and trend were not confidently identified, a standard and relatively simple configuration of (1,0,1) for the seasonality coefficient was chosen.

	(1,1,3)(0,0,0)	**(1,1,3)(1,0,1)**	(3,1,1)(0,0,0)	(3,1,1)(1,0,1)	(3,1,3)(0,0,0)	(3,1,3)(1,0,1)
AIC score	-337.46	**-340.05**	-333.27	-334.46	-334.84	**-339.32**
BIC score	**-320.59**	**-316.42**	-316.39	-310.83	-311.21	-308.95

Table 1: AIC and BIC values of ARIMA model combinations (best two in **bold**)

Following recommendations of Hyndman and Athanasopoulos (2018), the AIC and BIC score of the model combinations in Table 8 show the two best-performing models in each metric resulting in ARIMA(1,1,3)(1,0,1) as the best choice. When picking a configuration for an ARIMA model, paying attention to the residuals is also important. When looking at Figure 13, it is shown that the residuals do not follow a specific pattern, however, the histogram has two spikes that differ quite a lot compared to a normal distribution. Looking at the ACF plot, no lag shows outstanding significance. The p-value of the Ljung-Box test is between 0.05 and 0.01 and suggests that the residuals are not entirely independent. That means that the model might not have captured the underlying patterns of the data completely which might be a limiting factor on the overall performance of the model and should be kept in mind.

4.2 Exponential Smoothing (ETS)

The ETS model is commonly used to perform time series forecasts as a second option. Due to its ability to decompose the series into an error, trend, and seasonality component, it can capture the underlying patterns and use those to make predictions. Each component can be assumed to be either added or multiplied. Similarly to ARIMA, the most effective components of the ETS need to be found. The possible values for the error term include (A, M) for "additive" or "multiplicative". The trend can be assigned values in the range of (N, A, Ad, M, Md) meaning "None", "Additive", Additive damped",

”Multiplicative” and ”Multiplicative damped”) and the seasonal component can take on one of the values (N, A, M).

	AIC score	BIC score
ETS(A,N,A)	**226.43**	277.13
ETS(A,N,N)	**230.72**	**240.86**
ETS(A,A,N)	234.51	**251.41**
ETS(A,A,A)	232.53	289.99

Table 2: AIC and BIC values of ETS model combinations (best two in **bold**)

For the error term, the additive component makes the most sense seeing that the variability of the data is constant, as seen in Figure 4. The trend variable leaves more room for debate, as section 3 found that the trend is hard to be confidently identified. What can be said for certain, however, is that there is no exponentially increasing or decreasing trend, ruling out the multiplicative options for the parameter. Also, the need for damping is not present, as ongoing greenhouse gas emissions led to global temperatures rising over the past years and are not believed to naturally slow down at one point in the future. That leaves the option of selecting an additive trend or none at all. This paper will consider both options. For the seasonal component, the choice is similarly difficult, as only a very weak seasonal component was identified earlier in this paper. Again, the multiplicative option can be ruled out, leaving the additive option and not having a seasonal component. In order to get a notion of what combination would be performing best, the automatic ETS model is applied yielding the following result: ETS(A,N,A). Surprisingly it does not identify the need for a trend term. The performance of the models is compared in Table 2 by using the AIC and BIC score. ETS(A,N,N) is the most simple and simultaneously best-performing model and will be used for forecasting. An assessment of the residuals in Figure 12 shows that the residuals exhibit autocorrelation to some extent. Even if the line plot seems randomly distributed at first without a clear pattern, it still shows heteroscedasticity to some extent with a varying variance. Further, the distribution curve is not shaped like a normal distribution. Evaluating the ACF plots results in identifying one significant lag. The Ljung-Box test yields a value smaller than 0.05, thus the Null hypothesis is rejected, thereby contradicting independence between the residuals. This should be kept in mind when assessing the performance of the models.

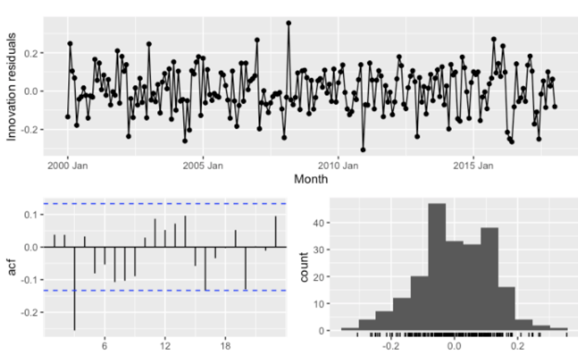

Figure 12: Residuals of ETS(A,N,N)

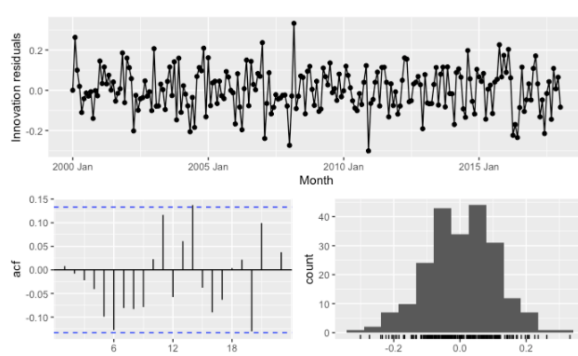

Figure 13: Residuals ARIMA(1,1,3)(1,0,1)

5 Results

The forecasts are displayed in Figure 14, with the actual development displayed in grey. The ARIMA model displays a slightly positive trend and exhibits a seasonal pattern that - purely visually - matches the ground truth better than the constant line in blue produced by the ETS(A,N,N). The ETS model shows no trend or seasonality, as both parameters are "None". Overall, both forecasts fall within the range of the ground truth but show less extreme swings in amplitudes, thus underestimating the actual data. As a rather simple model, ETS is obviously not able to capture the complexity of the data shown by its linearity and elliptic confidence intervals. The ARIMA model yields more precise results, given its narrower confidence intervals compared to the ETS model. The latter shows large confidence intervals, indicating more significant uncertainty in its model. The reason for that could be that the model is too simplistic or the data carries a complexity that is too high for the model to capture fully. To be able to assess the performance of the forecasts more quantitatively, Table 3 shows different metrics assessing the errors of the models, namely Root-mean-square-error (RMSE), mean-absolute-error (MAE), mean-absolute-percentage error (MAPE), and mean absolute scaled error (MASE). One can see that the ARIMA model has the lowest RMSE, MAE, MAPE, and MASE, concluding that it outperforms the ETS model. That confirms the notion based on the visual inspection of the forecasts. Thus this model is chosen to forecast until 2025. The result can be viewed in Figure 17. The prediction for the next three years exhibits a clear seasonal behavior, yet it appears to have a very weak or even non-existent trend. This is slightly surprising, as the trend component in Figure 4 is increasing. For the expected absolute values in that time period that means the surface temperature will be around 0.8 degrees Celcius warmer than the baseline period. While the predicted values still represent positive changes - thus a rise in temperature - the extent to which it rises is less than expected. Moreover, the prediction does not seem to capture the fluctuations of the previous years, showing much lower amplitudes. However, the large confidence intervals show that there might be way larger spikes in reality, which the model isn't entirely confident about.

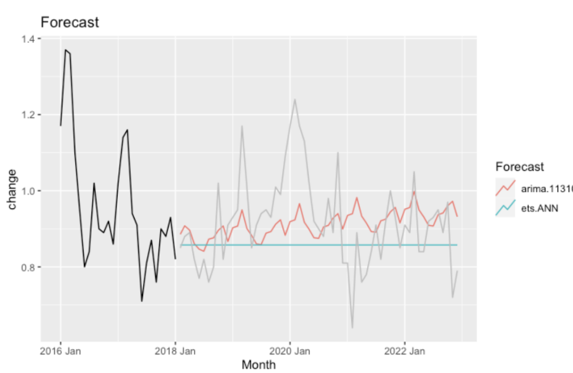

Figure 14: Forecasts for 2018-2022

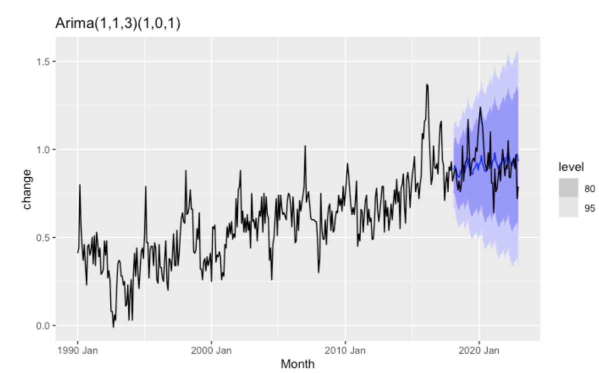

Figure 15: Forecast of ARIMA(1,1,3)(1,0,1)

Model	RMSE	MAE	MAPE	MASE
arima.113101	**0.116**	**0.089**	**9.774**	**0.605**
ets.ANN	0.132	0.099	10.270	0.675

Table 3: Metrics to assess accuracy of chosen models

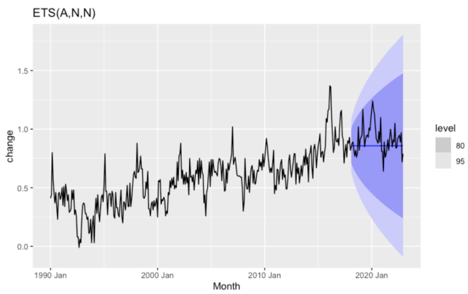

Figure 16: Forecast ETS(A,N,N)

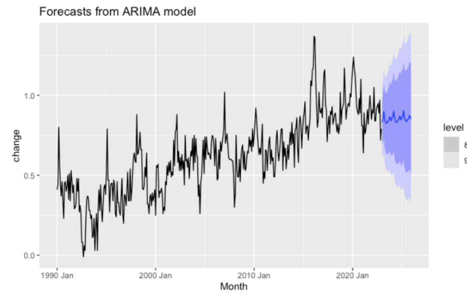

Figure 17: Forecasts for 2022-2025

6 Conclusion

This paper assessed the development of global surface temperature change grounded in the need to prevent misconceptions about climate change. Therefore multiple ETS and ARIMA models were tested to find the best configurations; in this case: ETS(A,N,N) and ARIMA(1,1,3)(1,0,1). Overall the ARIMA model performed best. A slight trend and seasonal component were detected during manual analysis but not confirmed by the automatic ARIMA and ETS models. The absence of trend and seasonality in the automatically selected models indicates a contradiction between the model's results and a perceived and publically discussed positive trend in global temperature throughout the last decades. Whereas an ETS(A,A,A) was tested in the attempt to include trend and seasonality in the forecast, it has not outperformed the models above and was thus not chosen for the forecast. Assessing whether there is no trend or if the model failed to detect it would be a suitable challenge for future work. Especially as the seemingly indisputable public perception of a clear rising trend gets challenged by statistical results, exploring this topic in future research seems highly interesting. The findings of the forecasts show that the average surface temperature is assumed to rise extremely slowly. The increase is assumed to be lower than intuitively expected based on the rise in the atmosphere temperature during the last years. If future work would confirm this, the disparity in the public perception of climate change and its actual development would be identified. A possible explanation for that disparity could be that the surface temperature is not as affected by climate change as the atmosphere and can thus not be compared well. Another possible cause could be that climate change leads to cold areas getting colder and warm areas getting warmer, thus balancing each other - something that the data set would not capture. Another possible limitation could be the lack of information about where exactly the data points were collected, as temperature variations change given the place and time. The models' performance is acceptable but still has room for improvement, especially given their large confidence intervals. Such improvements could include a larger train and test set, addressing potential over-differencing found in section 3.3 or reducing the autocorrelation between residuals. Further, monthly data might confuse the model with irregular fluctuations, while quarterly data could be more suitable for a simple model like ETS(A,N,N). To conclude, this paper predicted the change in global surface temperature to be on a very similar level three years from now, around +0.8 degrees Celcius, deviating from the widespread idea of rapidly escalating temperature changes. While this should not be understood as a denial of climate change, it does indicate that the pace and extent of these changes might be more gradual than is often presented in public debate. This paper can be built upon to perform further statistical exploration and more accurate predictions regarding climate change.

References

J. Hansen, R. Ruedy, M. Sato, and K. Lo. Global surface temperature change. *Rev. Geophys.*, 48:RG4004, 2010. doi: 10.1029/2010RG000345.

J. Hansen, R. Ruedy, M. Sato, and K. Lo. GISS Surface Temperature Analysis (GISTEMP v4). `https://data.giss.nasa.gov/gistemp/tabledata`$_v 4/GLB.Ts + dSST.csv$, 2023. $Accessed : 2023 - 06 - 01$.

R. J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2018.

Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. 2020. URL `https://otexts.com/fpp3/`.

N. Lenssen, G. Schmidt, J. Hansen, M. Menne, A. Persin, R. Ruedy, and D. Zyss. Improvements in the gistemp uncertainty model. *Journal of Geophysical Research: Atmospheres*, 124(12):6307–6326, 2019. doi: 10.1029/2018JD029522.

Florian Oswald, Vincent Viers, Pierre Villedieu, and Gustave Kennedi. *onstationarity II: Breaks*. SciencesPo Department of Economics, 2020. Accessed: 2023-06-16.

A Appendix

A.1 ADF and KPSS

```
[1] "ADF with trend-----------------------------------"

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-0.31919 -0.07205  0.00210  0.07120  0.41324

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.581e-02  9.836e-03  -1.608   0.1084
z.lag.1     -2.571e-01  3.710e-02  -6.931 1.10e-11 ***
tt           4.089e-04  6.386e-05   6.404 3.09e-10 ***
z.diff.lag1 -2.551e-01  4.590e-02  -5.559 4.11e-08 ***
z.diff.lag2 -6.215e-03  4.588e-02  -0.135   0.8923
z.diff.lag3 -9.033e-02  4.087e-02  -2.210   0.0275 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1068 on 593 degrees of freedom
Multiple R-squared:  0.2435,    Adjusted R-squared:  0.2372
F-statistic: 38.18 on 5 and 593 DF,  p-value: < 2.2e-16


Value of test-statistic is: -6.9306 16.0629 24.021

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Figure 18: ADF with "trend"

```
[1] "ADF with drift-----------------------------------"

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-0.29778 -0.06880 -0.00193  0.06809  0.38692

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.012686   0.008328   1.523 0.128252
z.lag.1      -0.016732   0.015552  -1.076 0.282446
z.diff.lag1  -0.474357   0.043572 -10.887  < 2e-16 ***
z.diff.lag2  -0.202343   0.047724  -4.240 2.61e-05 ***
z.diff.lag3  -0.270987   0.048349  -5.605 3.25e-08 ***
z.diff.lag4  -0.138756   0.049571  -2.799 0.005297 **
z.diff.lag5  -0.141107   0.049616  -2.844 0.004615 **
z.diff.lag6  -0.144988   0.049516  -2.928 0.003546 **
z.diff.lag7  -0.168545   0.049719  -3.390 0.000747 ***
z.diff.lag8  -0.123722   0.049802  -2.484 0.013267 *
z.diff.lag9  -0.133119   0.049792  -2.674 0.007720 **
z.diff.lag10 -0.136812   0.049719  -2.752 0.006116 **
z.diff.lag11 -0.058976   0.049527  -1.191 0.234230
z.diff.lag12 -0.058995   0.049442  -1.193 0.233286
z.diff.lag13 -0.061413   0.049428  -1.242 0.214569
z.diff.lag14 -0.062719   0.049395  -1.270 0.204690
z.diff.lag15 -0.112400   0.049384  -2.276 0.023212 *
z.diff.lag16 -0.119541   0.049269  -2.426 0.015561 *
z.diff.lag17 -0.137033   0.049179  -2.786 0.005506 **
z.diff.lag18 -0.136239   0.049135  -2.773 0.005740 **
z.diff.lag19 -0.060468   0.048828  -1.238 0.216082
z.diff.lag20 -0.132242   0.048470  -2.728 0.006561 **
z.diff.lag21 -0.118276   0.048381  -2.445 0.014799 *
z.diff.lag22 -0.035421   0.048156  -0.736 0.462319
z.diff.lag23 -0.048437   0.046815  -1.035 0.301266
z.diff.lag24  0.090451   0.046046   1.964 0.049971 *
z.diff.lag25  0.108676   0.041757   2.603 0.009493 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1076 on 572 degrees of freedom
Multiple R-squared:  0.2596,    Adjusted R-squared:  0.2259
F-statistic: 7.714 on 26 and 572 DF,  p-value: < 2.2e-16


Value of test-statistic is: -1.0759 1.2338

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

Figure 19: ADF with "drift"

```
[1] "ADF with none----------------------------------"

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-0.29102 -0.06575  0.00280  0.07597  0.39283

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
z.lag.1       0.003219   0.008396   0.383 0.701565
z.diff.lag1  -0.489904   0.042408 -11.552  < 2e-16 ***
z.diff.lag2  -0.215627   0.046975  -4.590 5.44e-06 ***
z.diff.lag3  -0.283264   0.047728  -5.935 5.09e-09 ***
z.diff.lag4  -0.149718   0.049103  -3.049 0.002402 **
z.diff.lag5  -0.151419   0.049209  -3.077 0.002190 **
z.diff.lag6  -0.154403   0.049186  -3.139 0.001781 **
z.diff.lag7  -0.177120   0.049457  -3.581 0.000371 ***
z.diff.lag8  -0.131363   0.049606  -2.648 0.008317 **
z.diff.lag9  -0.140138   0.049635  -2.823 0.004917 **
z.diff.lag10 -0.143016   0.049609  -2.883 0.004088 **
z.diff.lag11 -0.064121   0.049468  -1.296 0.195432
z.diff.lag12 -0.063647   0.049405  -1.288 0.198169
z.diff.lag13 -0.065899   0.049397  -1.334 0.182707
z.diff.lag14 -0.066965   0.049373  -1.356 0.175534
z.diff.lag15 -0.116267   0.049375  -2.355 0.018871 *
z.diff.lag16 -0.122774   0.049280  -2.491 0.013007 *
z.diff.lag17 -0.139724   0.049204  -2.840 0.004676 **
z.diff.lag18 -0.138080   0.049177  -2.808 0.005158 **
z.diff.lag19 -0.061432   0.048880  -1.257 0.209343
z.diff.lag20 -0.133011   0.048523  -2.741 0.006313 **
z.diff.lag21 -0.118411   0.048437  -2.445 0.014799 *
z.diff.lag22 -0.034749   0.048210  -0.721 0.471333
z.diff.lag23 -0.047700   0.046866  -1.018 0.309211
z.diff.lag24  0.091204   0.046096   1.979 0.048344 *
z.diff.lag25  0.109046   0.041804   2.608 0.009332 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1077 on 573 degrees of freedom
Multiple R-squared:  0.2566,	Adjusted R-squared:  0.2229
F-statistic: 7.608 on 26 and 573 DF,  p-value: < 2.2e-16


Value of test-statistic is: 0.3834

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Figure 20: ADF with "none"

```
[1] "KPSS with tau----------------------------------"

#######################
# KPSS Unit Root Test #
#######################

Test is of type: tau with 6 lags.

Value of test-statistic is: 0.1137

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216

[1] "KPSS with mu----------------------------------"

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 8.0191

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Figure 21: KPSS tests

A.2 EDA

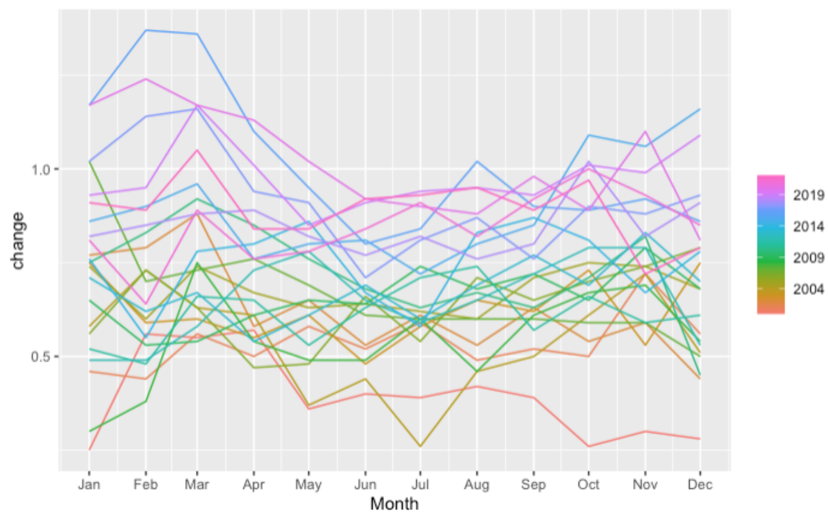

Figure 22: Temperature changes over months 2000-2022